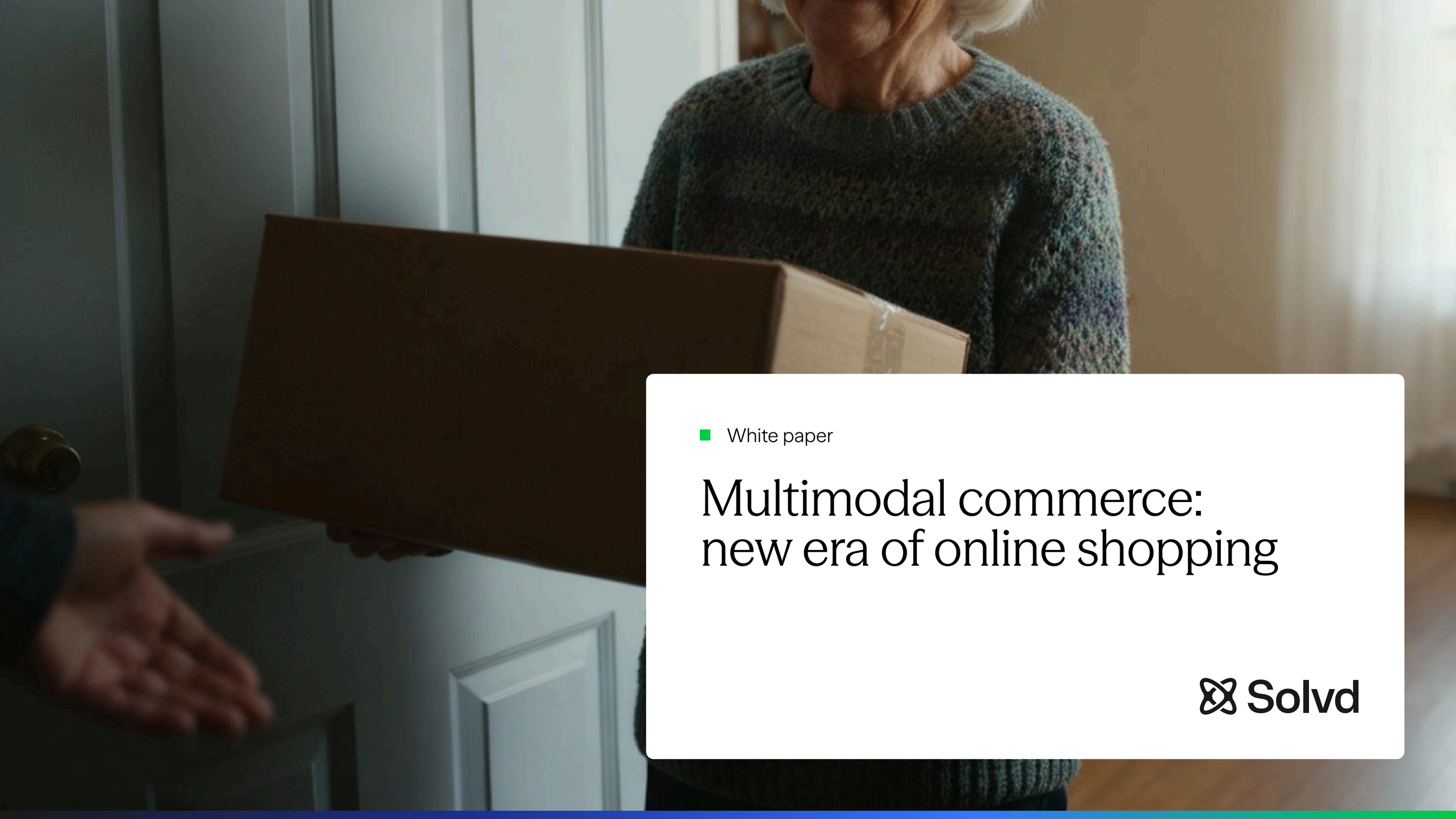


White paper

Multimodal commerce: new era of online shopping

Solvd

Contents

Solvd

Executive summary

Humans are multimodal. Sight, hearing, smell, taste and touch help us perceive different types of information to explore the world and make context-rich decisions. Today, the same experience has become possible with multimodal commerce which is transforming the retail industry and opening new horizons for customers and brands alike. Multimodal commerce has multimodal AI at its core, which refers to an AI system that is able to process and integrate information from multiple types of input data, such as text, voice, image, video and AR. This ability enables companies to create and deliver personalized, connected journeys wherever customers engage with their brand.

This white paper explores the definition of multimodal commerce and multimodal AI, as well as adoption statistics, benefits and real-world use cases that reveal the hands-on experiences of early adopters.

Who is it for?

Retail and e-commerce executives, product and innovation teams, technology leaders and AI practitioners.

Why it matters

53%

of consumers switch brands regularly, despite subscribing to their loyalty programs. Experimentation and lack of personalization are major reasons for switching[1]

28%

of consumers said they'd used AI-powered visual search to find products that match or resemble items they want to buy[2]

2/3

of Gen Z and millennials want hyper-personalized content and product recommendations, powered by Gen AI[1]

Current adoption statistics and market growth

Modern buying journeys illustrate why customer experience is now more essential than ever before. Today's buyer might browse on mobile, research in ChatGPT, call support and make a purchase in-store, all within a 24-hour period. But when these touchpoints don't "talk" to each other, when they feel like separate, independent components, it results in frustration, abandonment and lost loyalty. Multimodal commerce solves this by unifying these interactions into one coherent, personalized and context-aware journey.

40%
of generative AI solutions will be multimodal (text, image, audio and video) by 2027, up from 1% in 2023[3]

71%
of consumers want generative AI integrated into their shopping experiences[1]

15-20%
is an average cost savings for companies using multimodal transportation compared to unimodal solutions[5]

3-4x
higher likelihood that shoppers who interact with chatbots will convert[6]

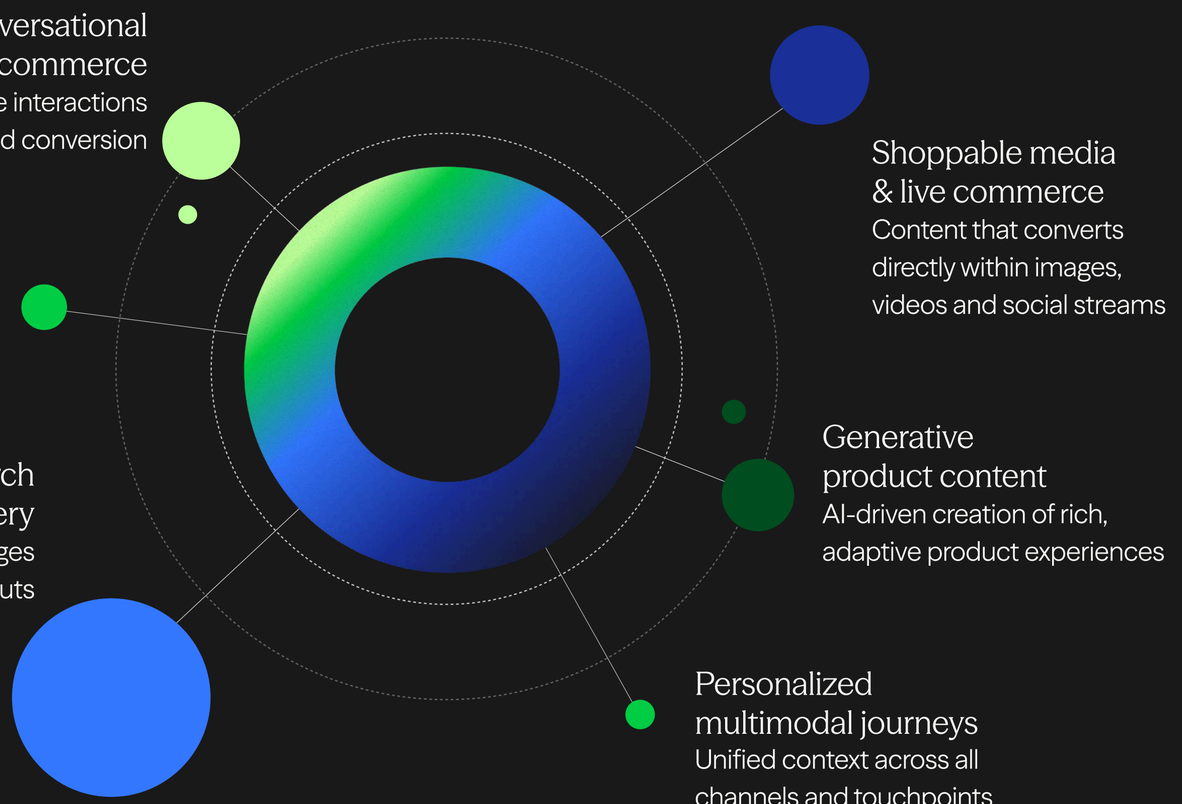
$1.6B
is the global multimodal AI market size, and it is estimated to grow at CAGR of 32.7% from 2025 to 2034[4]

78%
of companies that implemented multimodal solutions reported improved delivery times[5]

16%
of multimodal AI deployments are focusing on enhancing customer experience and personalization[7]

What is multimodal commerce?

Multimodal commerce is the next generation of online shopping, enabling customers to discover, interact with and buy products seamlessly through every possible type of interaction. Multimodal commerce has multimodal AI at the core, which refers to AI system that is able to process and integrate information from multiple types of input data, such as text, voice, image, video and AR. This ability enables companies to create and deliver personalized, connected journeys wherever customers engage with their brand, building stronger customer loyalty and driving repeat purchases and cross-selling.

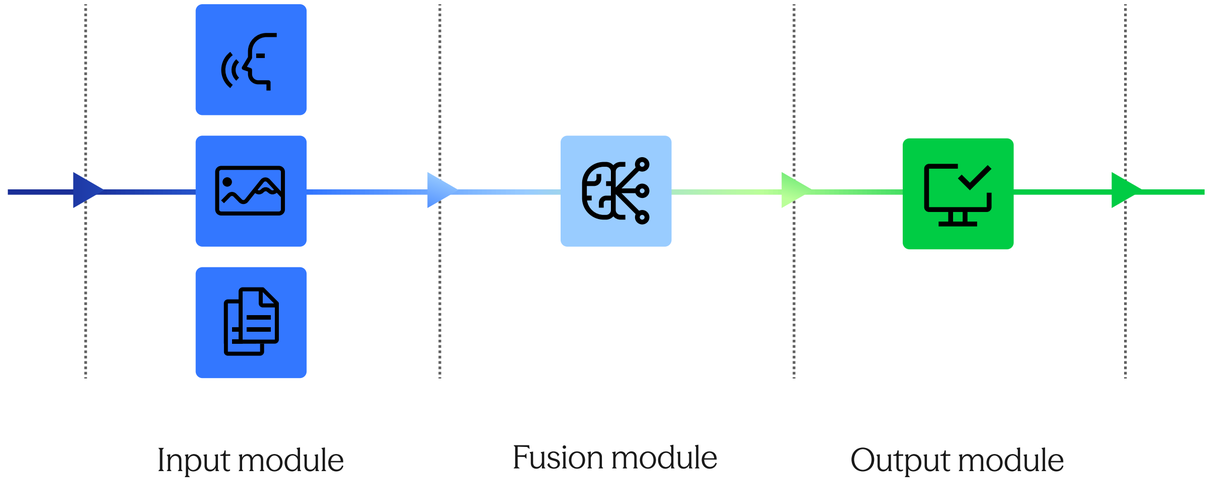


How does it work?

A multimodal AI system begins by translating each input type - text, images, audio or video - into a standardized representation. These signals are then fused inside a shared intelligence layer that understands how they relate to each other. From this unified understanding, the model can generate text, interpret visuals, answer spoken questions or combine modalities seamlessly, delivering a more natural and context-aware customer experience.

A multimodal AI system works like a skilled human expert who listens, looks and reads at the same time - combining all inputs to form a clearer, more accurate understanding before responding.

Three key components of multimodal AI



Multimodal commerce capabilities

The table below outlines each capability of multimodal commerce and its positive effect on the most important aspects of retail.

		Consumer experience	Consumer acquisition	Revenue/ basket size	Operational cost
Cell phone product scanner	Turns phones into instant product scanners with auto-pricing, listing drafts and AR overlays.	■		■	■
Virtual try-on	Enables virtual try-ons so shoppers can see how clothes fit on real bodies before buying.	■	■	■	
Style add-ons	Delivers outfit inspiration features that suggest complementary items and boost basket size.	■	■	■	
Listing creators	Automated listing creation with one-click AI that generates images, titles and descriptions.	■	■		■
Digital marketing optimization	Improved communication performance using AI-optimized subject lines for higher open rates.		■	■	
Upgraded product photography	Generates professional model photography from plain product shots using AI.	■			■
Counterfeit detection	Builds visual authentication tools to detect counterfeits by comparing return and original item images.				■
Advanced product attributes	Enhanced image understanding capabilities to extract product attributes and improve search relevance.	■			
Smart search	Boosts search accuracy with LLM-driven ranking aligned to shopper intent.	■		■	

A comparative overview of core operational aspects

To better understand the difference between multimodal and unimodal AI systems, let's take a look at the table, which outlines their distinctions across core dimensions including definitions, examples, technical capabilities and business value.

Multimodal AI systems vs. unimodal AI systems

Multimodal AI systems		Unimodal AI systems
Combine text, voice, image, video, gesture and other sensory signals	Input types	Accept a single input type (e.g., only natural language text)
Produce rich, cross-modal responses, such as generating visuals from text or responding to spoken queries	Output capability	Produce output only in the same modality as the input (e.g., text in → text out)
Build a more holistic understanding and context by analyzing what is said and what is shown	Understanding & context	Limited understanding based on one data source which may lead to missing contextual clues available in other modalities
More accurate and resilient. One modality can compensate for another's ambiguity	Accuracy & robustness	Might be less accurate in ambiguous or noisy scenarios
Require cross-modal fusion, alignment and synchronization between heterogeneous data sources	Technical complexity	Simpler architecture and training as the system focuses on a single data type
Deliver richer personalization, higher engagement and more human-like commerce experiences	Business value	Deliver focused, functional experiences but limited in engagement potential
A user sends an image of shoes and asks for "similar ones under $100" and the system combines vision and language	Use-case example in retail	A chatbot recommends shoes based only on text prompts

Real-world use cases

Part 1

Today, the benefits of multimodal AI in retail are real, measurable and even publicly available. Together, these examples highlight how multimodal AI is being applied across inventory, logistics, customer support and product discovery to unlock scale and competitive advantage.

JD Sports, AI-enabled logistics[8]

JD Sports, a sports-fashion retail leader, leverages massive AI-enabled logistics and warehousing operations. Empowered by computer vision, robotics and prediction models, their supply chain is able to manage over 10 million SKUs and maintain an industry-leading inventory turnover of ~30 days while still offering lightning-fast delivery. For 2024, AI-driven optimization of inventory replacement and routing enables JD Sports to deliver 95% of their orders within 24 hours across China. They also developed a time-series forecasting AI (TimeHF) that improved the accuracy of demand forecasting by over 10% through ensuring products are stocked in the right quantities and locations and are always available for customers. In terms of business outcomes, this approach helped JD Sports decrease the fulfillment expense rate to 5.8% of revenue, which indicates cost efficiency and sales growth without inventory issues[9].

H&M, AI chatbot for customer support[10]

H&M, a global fashion retailer operating in markets including the United States and China, uses an AI-powered chatbot on its websites and messaging channels to help customers cope with product queries, sizing and order issues. Now, all of these issues can be solved through an automated conversational system that guides customers step-by-step and provides instant answers across both markets. The company and its technology partners have reported that the chatbot significantly improves response speed; case studies note reductions of up to 70% compared with human-handled inquiries. This led to a reduction in the workload on support staff and faster response times[11].

Home Depot, a computer-vision “Sidekick” application[12]

Home Depot, an American multinational home improvement corporation, equipped store associates with a computer-vision “Sidekick” application to monitor and improve on-shelf availability, which removes friction for both employees and customers. By combining image input and inventory data, this multimodal approach enables employees to snap photos of overhead storage shelves and quickly gain insights into what items are stored high, which need restocking on the sales floor or replenish before shelves go empty. As a result, the deployment of this application helps Home Depot to free up time and focus on customer assistance instead of scanning labels, detect misplaced items and increase e-commerce order picking from stores[13].

Real-world use cases

Part 2

Today, the benefits of multimodal AI in retail are real, measurable and even publicly available. Together, these examples highlight how multimodal AI is being applied across inventory, logistics, customer support and product discovery to unlock scale and competitive advantage.

Alibaba's Taobao, image-based product search[14]

Alibaba's Taobao, China's largest e-commerce marketplace, enabled image-based product search that lets customers upload photos to Taobao in order to find the same or similar products. In mid-2025, Taobao updated its AI-based image search system by introducing support for higher resolution and multi-angle recognition. This approach breaks language barriers and simplifies discovery[15]. Although Alibaba does not publicly disclose conversion rates for Taobao's visual search, consumer behavior indicates its value: a 2023 study found that 62% of Gen Z and millennials are more likely to purchase if they can search for products by image[6].

Amazon, AI-powered visual search[16]

Amazon, an American multinational technology company, empowered its retail application with StyleSnap, which allows users to upload a photo of an outfit or home item and receive matching product recommendations. This visual discovery approach dramatically shortens the path from inspiration to purchase. According to Gartner, early adopters of visual and voice searches, such as Amazon, could increase digital commerce revenue by up to 30%, while traditional search engine volume will drop by 25%[17, 18].

Pinterest, visual search and AR try-on[19]

Pinterest, an American social media service with over 570 million monthly users, noted that visual discovery generates a strong ROI in retail advertising. Their visual-search systems (like Lens or "Shop the Look") use computer vision and deep embeddings to allow users to find products directly from images. Retailers on Pinterest saw a 73.9% increase in conversion and 430% return on ad spend from visual search-based ads (e.g. Lens) turning inspiration images into purchases. While 80% of users say that the main reason for using the platform is inspiration, more than 85% have made a purchase based on Pins they saw from brands[6, 20].

Conclusion

Multimodal commerce offers an unprecedented shift in the way retailers operate and customers make purchases. Seamless omnichannel interactions, personalized experiences, enhanced customer support, better demand forecasting and richer customer insights — all of these benefits make the shopping experience more convenient, intuitive and responsive to constantly evolving consumer expectations.

One of the core advantages of multimodal commerce is enhanced accessibility. By going beyond a single data or interaction type, multimodal AI systems create a more inclusive shopping environment and support customers with different abilities, preferences and device constraints.

As multimodal technologies continue to mature, the retailers who invest in these technologies today will
meet customers' high expectations and define the competitive landscape of tomorrow.

Take the first step into the new era of the shopping experience

About Solvd

Solvd is an AI-first advisory and digital engineering firm helping enterprises bridge the gap between experimentation and execution, delivering real ROI. Our unique capabilities combine deep implementation experience with world-class academic research, supported by contributions to leading conferences such as NeurIPS, ICML and ECCV. With expertise across AI advisory, data engineering, digital experience, application development, cloud engineering and quality engineering & GRC, Solvd empowers enterprises in healthcare, life sciences, media, retail and beyond to turn AI potential into production-ready solutions and measurable business impact.

Visit us at Solvd.com

Solvd

Bibliography

1. Capgemini, 2025. Research Institute's annual consumer trends report 'What Matters to Today's Consumer'. https://www.capgemini.com/news/press-releases/71-of-consumers-want-generative-ai-integrated-into-their-shopping-experiences/

2. BCG, 2024. Consumers Know More About AI than Business Leaders Think. https://www.bcg.com/publications/2024/consumers-know-more-about-ai-than-businesses-think

3. Gartner, 2024. What Generative AI Means for Business. https://www.gartner.com/en/insights/generative-ai-for-business

4. Global Market Insights, 2025. Multimodal AI Market Size. https://www.gminsights.com/industry-analysis/multimodal-ai-market

5. Gitnux, 2025. Multimodal Statistics. https://gitnux.org/multimodal-statistics/

6. Envive, 2025. 52 Online Shopping Conversion Lift Statistics in 2025. https://www.envive.ai/post/online-shopping-conversion-lift-statistics

7. Zebracat, 2025. 50+ Multimodal AI Market Size Insights And Growth Projection. https://www.zebracat.ai/post/multimodal-ai-market?

8. JD Sports, 2025. Official website. https://www.jdsports.com/

9. JD Sports, 2023. Corporate blog. https://jdcorporateblog.com/pioneering-digital-transformation-jd-coms-robust-presence-at-ciftis-2023/

10. H&M, 2025. Official website. https://www2.hm.com/

11. Sama, 2023. 5 Computer Vision Applications in Retail in 2023. https://www.sama.com/blog/computer-vision-applications-in-retail

12. Home Depot, 2025. Official website. https://www.homedepot.com

13. Home Depot, 2024. Annual report 2024. https://ir.homedepot.com/~/media/Files/H/HomeDepot-IR/2025/HD_2024_AR_IRsite_v2.pdf

14. Taobao, 2025. Official website. https://www.taobao.com/

15. The Chronicle Journal, 2025. New Taobao Image Search Update Enhances Reverse Image Shopping for Global Dropshippers.https://markets.chroniclejournal.com/chroniclejournal/article/abnewswire-2025-7-16-new-taobao-image-search-update-enhances-reverse-image-shopping-for-global-dropshippers

16. Amazon Web Services, Inc., 2025. Official website. https://aws.amazon.com/

17. Netguru, 2025. 5 Ways Computer Vision Is Transforming Online Shopping. https://www.netguru.com/blog/computer-vision-online-shopping

18. Gartner, 2024. Gartner Predicts Search Engine Volume Will Drop 25% by 2026, Due to AI Chatbots and Other Virtual Agents. https://www.gartner.com/en/newsroom/press-releases/2024-02-19-gartner-predicts-search-engine-volume-will-drop-25-percent-by-2026-due-to-ai-chatbots-and-other-virtual-agents

19. Pinterest, 2025. Official website. https://www.pinterest.com/

20. Neil Patel, 2025. A Beginner's Guide to Pinterest Ads. https://neilpatel.com/blog/pinterest-ads/